

A Fast Approximation for Adaptive Wavelength Selection for Infrared Chemical Sensors[†]

Mark Chilenski,* Cara Murphy, and Gil Raz

Systems & Technology Research, Woburn, MA, USA

E-mail: mark-DOT-chilenski-AT-stresearch-DOT-com

Abstract

Active mid-infrared spectroscopy with tunable lasers is a leading technology for standoff detection and identification of trace chemicals. Information-theoretic optimal selection of the laser wavelength offers the promise of increased detection confidence at lower abundances and with fewer wavelengths. Reducing the number of wavelengths required enables faster detections and lowers sensor power consumption while keeping the optical power under eye safety limits. This paper presents an approximation to the mutual information which operates $\sim 40\,000\times$ faster than traditional techniques, thereby making near-optimal real-time sensor control computationally feasible. Application of this technique to synthetic data suggests it can reduce the number of wavelengths needed by a factor of two relative to an evenly-spaced grid, with even higher gains for chemicals with weak signatures.

Keywords

chemical identification, active mid-IR spectroscopy, hyperspectral imaging, sensor control, information theory

[†]Preprint to be submitted to *ACS Sensors*. © 2018 Systems & Technology Research.

Active mid-infrared (MIR) spectroscopy¹ is a popular technique for detection and identification of trace chemicals (both surface residues and vapors/aerosols) at distances of up to tens of meters.^{2–13} The MIR reflectance spectrum can be obtained either by illuminating the target with a broadband source and dispersing the received light at the detector, or by scanning the wavelength¹⁴ of a tunable, narrowband source such as a quantum cascade laser.^{10,11} Use of a tunable source offers many advantages, including higher optical throughput (for a given level of illumination, such as dictated by eye safety), simpler detectors (no need for a Fourier-transform infrared or other complex spectrometer), and the ability to select only wavelengths which help discriminate between target chemicals (i.e., no photons need be emitted at uninformative wavelengths).

A block diagram of a laser-based chemical sensor is shown in Figure 1. The active IR

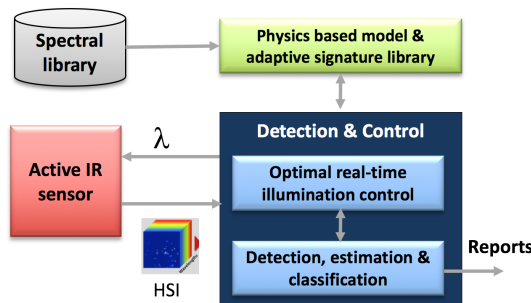


Figure 1: Block diagram of an active MIR chemical sensor. An active IR sensor interrogates the target with a tunable laser and reports the reflectance spectrum to the detection and control algorithm. The detection algorithm compares the measured spectrum to a library of spectral signatures, and the control algorithm selects additional wavelengths to measure until the desired level of confidence is reached.

sensor comprises a tunable laser (which can be raster-scanned over the target to build up an image) and a broadband imager. The active sensor reports the reflectance spectrum, in the form of a hyperspectral image (HSI), to the detection and control algorithm. The detection algorithm^{15–18} compares the measured spectrum to a library of spectral signatures. If the results of the detection algorithm are inconclusive, the control algorithm selects further wavelengths to measure. This process repeats until the desired detection confidence is obtained.

Information theory is often used for feature selection in the machine learning community,¹⁹ and has been applied to the optimization of various chemical sensors,^{20–22} the characterization of spectral variability in hyperspectral images,²³ and the characterization of limits of detection in chemical sensors.²⁴ In particular, previous work has presented techniques for information-theoretic *optimal wavelength selection* (OWLS).²⁵ Specifically, given candidate wavelengths Ω , OWLS seeks

$$\Lambda = \arg \max_{\Lambda \subset \Omega, |\Lambda| \leq k} I(R(\Lambda); Y), \quad (1)$$

where $R(\Lambda)$ is the reflectance at the wavelengths Λ , Y is the identity of the chemical, and $I(R; Y)$ is the mutual information between R and Y .²⁶ The wavelengths selected with this scheme are optimal in the sense that maximizing $I(R; Y)$ minimizes the mis-classification rate.²⁷

The simplest way to use OWLS is *a priori*: the sequence of wavelengths which is expected to deliver the best accuracy is selected offline, before any data have been measured. It is desirable, however, to select the next wavelength(s) to measure adaptively based on the data which have been measured so far, as described in Algorithm 1.

For offline use, $I(R; Y)$ is estimated at each step (line 6) using the Kozachenko-Leonenko (KL) estimator.²⁸ Computing the first 10 wavelengths (from 500 candidates) to optimally discriminate between 67 chemicals takes six minutes on a typical laptop, and is therefore infeasible for real-time, adaptive wavelength selection. Instead, this paper presents a geometric approximation which exploits the structure of the signature model to deliver comparable results in under 10 ms: a $40\,000\times$ speedup.

1 Description of the Algorithm

In order to enable real-time adaptive OWLS, we need an algorithm which has the following properties:

Algorithm 1 Adaptive OWLS (Algorithm 2 of Ref. 25)

```
1: function ADAPTIVEOWLSDETECTION( $\Lambda, \Omega, \alpha, q$ ) ▷  $\Lambda$  is the initial set of  
   wavelengths ▷  $\Omega$  is the set of candidate wavelengths ( $\Lambda \cap \Omega = \emptyset$ )  
   ▷  $\alpha$  is the required level of detection confidence  
   ▷  $q$  is the quantile of chemicals to retain at each step  
2:    $\mathcal{D} \leftarrow \text{MEASURE}(\Lambda)$  ▷  $\mathcal{D}$  is the measured data at the wavelengths in  $\Lambda$   
3:    $c, s \leftarrow \text{DETECT}(\mathcal{D})$  ▷  $c$  is the confidence that the chemical with the highest score is  
   correct ▷  $s$  is the detection score for each candidate chemical  
4:   while  $c < \alpha$  and  $|\Omega| > 0$  do  
5:      $\hat{Y} \leftarrow \{y \mid s_y \geq q(s)\}$  ▷  $\hat{Y}$  are the most likely chemicals  
6:      $\lambda_{\text{new}} \leftarrow \arg \max_{\lambda \in \Omega} I(R(\Lambda \cup \{\lambda\}); \hat{Y})$  ▷ Add wavelength(s) which best distinguish  
   between  $\hat{Y}$   
7:      $\Lambda \leftarrow \Lambda \cup \{\lambda_{\text{new}}\}$   
8:      $\Omega \leftarrow \Omega \setminus \{\lambda_{\text{new}}\}$   
9:      $\mathcal{D} \leftarrow \mathcal{D} \cup \text{MEASURE}(\lambda_{\text{new}})$   
10:     $c, s \leftarrow \text{DETECT}(\mathcal{D})$   
11:  return  $\arg \max s$  ▷ Return chemical with highest detection score
```

- Selects wavelengths which contain comparable mutual information to those selected by more rigorous means (such as the KL estimator).
- Can incrementally add new wavelengths to the existing set Λ .
- Can select wavelengths in less time than it would take to simply measure all of the candidate wavelengths. (Current systems can measure as fast as 1 ms per wavelength.¹³)

The geometric approximation is inspired by the structure of the signature model illustrated in Figure 2: the cluster corresponding to each chemical forms a line emanating from a single point (the reflectance of the bare substrate). For double pass absorption (such as from a thin film on a metallic surface), the linear relationship is exact in the absorbance domain (i.e., the logarithm of reflectance). For more general cases, this linear behavior is a reasonable approximation for low abundances (where the signal-to-noise ratio (SNR) is low and the most benefit can be gleaned from OWLS). Therefore, each chemical can be thought of as corresponding to a vector which begins at the point corresponding to bare substrate

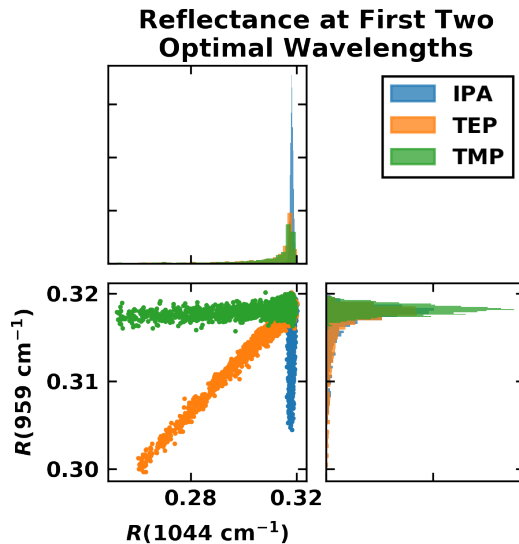


Figure 2: Joint distribution of reflectance at the first two optimal wavelengths to classify between isopropyl alcohol (IPA), triethyl phosphate (TEP), and trimethyl phosphate (TMP). The lower left panel is the bivariate joint distribution and the panels on the diagonal are the univariate marginal distributions. The points represent various abundances and realizations of random noise.

and ends at the point corresponding to a typical abundance. A possible way to pick informative wavelengths is then to find the set Λ which maximizes the squared Euclidean distance between the endpoints of all pairs of chemicals a and b :

$$D_{a,b}^2(\Lambda) = \sum_{\lambda \in \Lambda} (R_a(\lambda) - R_b(\lambda))^2, \quad (2)$$

where $R_a(\lambda)$ is the reflectance at wavelength λ for a typical concentration of chemical a . Simply picking wavelengths which maximize the sum of this quantity over all pairs of chemicals will not provide a good approximation to the wavelengths selected using an actual calculation of mutual information, however: given an initial set of wavelengths, the next wavelength should be the one which contributes the most to separating the chemical pairs *which are not already well-separated*. Therefore, when assessing the potential gain from adding a wavelength, the objective function should include some form of discounting to reduce the weight of gains which come from chemical pairs which are already well-separated. This suggests a

function of the form

$$G(\lambda) = \left(\sum_{\substack{a,b \\ b>a}} \left| \frac{D_{a,b}^2(\lambda)}{D_{a,b}^{2\delta}(\Lambda)} \right|^p \right)^{1/p}, \quad (3)$$

which has two parameters:

- $\delta \geq 0$ sets how strongly gains for chemical pairs which already have large separation are discounted. Setting $\delta = 0$ ignores how well the current set Λ already separates any given pair of chemicals.
- p selects how the discounted gains for each chemical pair are aggregated. The extremes are $p = 1$ (sum of discounted gains) and $p = \infty$ (maximum of discounted gains), with intermediate values interpolating between the two behaviors.

The full procedure for geometric real-time optimal wavelength selection (GROWLS) is shown in Algorithm 2. In practice, this function replaces the computation of $I(R; Y)$ on line 6 of Algorithm 1.

2 Parameter Tuning

In order for GROWLS to be effective, the parameters δ and p must be tuned so that it best approximates the behavior of the KL estimator. One way of selecting the parameters *a priori* is to minimize the sum of the difference in mutual information captured by the KL estimator and GROWLS:

$$\Delta = \sum_{i=1}^k \left(I(R(\{\lambda_{\text{KL},1}, \dots, \lambda_{\text{KL},i}\}); Y) - I(R(\{\lambda_{\text{geom},1}, \dots, \lambda_{\text{geom},i}\}); Y) \right), \quad (4)$$

where $k = |\Lambda|$ is the number of wavelengths selected. Figure 3(a) shows Δ as a function of δ and p for $k = 50$. For very low δ , too many wavelengths which only help easily-separable

Algorithm 2 Geometric Real-Time Optimal Wavelength Selection (GROWLS)

1: **function** GROWLS(Ω, δ, p, k)

 $\triangleright \Omega$ is the set of candidate wavelengths

 $\triangleright \delta$ is the discount exponent

 $\triangleright p$ is the norm order

 $\triangleright k \leq |\Omega|$ is the number of wavelengths to select

2: $\lambda_{\text{new}} \leftarrow \arg \max_{\lambda \in \Omega} \sum_{\substack{a,b \\ b > a}} D_{a,b}^2(\lambda)$

3: $\Lambda \leftarrow \{\lambda_{\text{new}}\}$

4: $\Omega \leftarrow \Omega \setminus \{\lambda_{\text{new}}\}$

5: **while** $|\Lambda| < k$ **do**

6: $\lambda_{\text{new}} \leftarrow \arg \max_{\lambda \in \Omega} \left(\sum_{\substack{a,b \\ b > a}} \left| \frac{D_{a,b}^2(\lambda)}{D_{a,b}^{2\delta}(\Lambda)} \right|^p \right)^{1/p}$

7: $\Lambda \leftarrow \Lambda \cup \{\lambda_{\text{new}}\}$

8: $\Omega \leftarrow \Omega \setminus \{\lambda_{\text{new}}\}$

9: **return** Λ

chemical pairs are selected, which results in performance which is worse than using evenly-spaced wavenumbers. For all three values of p there is a broad minimum near $\delta = 1.6$. There is only a weak dependence on p , but the minimum for $p = 1$ is slightly lower than for $p = 2$ and $p = \infty$. Therefore, the parameters $p = 1$, $\delta = 1.6$ were used for the remainder of this work.

Figure 3(b) shows the trajectories of mutual information accumulation for evenly-spaced wavenumbers, the optimal wavelengths computed using the KL estimator, and the approximate wavelengths computed using GROWLS. The geometric approximation nearly matches the performance of the KL estimator, indicating that this scheme should deliver comparable detection performance to the full mutual information calculation. Figure 3(c) shows the spectral library of 67 chemicals together with the wavelengths selected by the three techniques. Both the KL estimator and GROWLS pick wavelengths which are focused on informative parts of the spectrum, ignoring the uninformative regions that the evenly-spaced scheme samples.

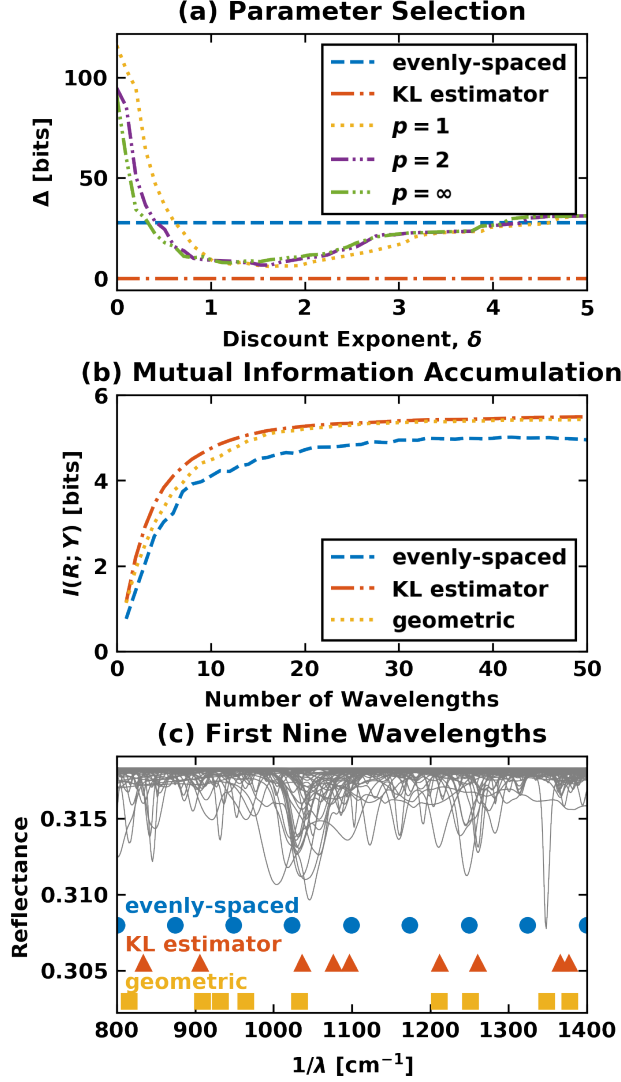


Figure 3: (a) Summed mutual information difference Δ for $k = 50$ wavelengths as δ and p are varied. There is a broad minimum for $\delta \approx 1.6$ and a fairly weak dependence on p . Also shown is the mutual information difference when using evenly-spaced wavenumbers (blue dashed line) and the bound of $\Delta = 0$ corresponding to exact reproduction of the results of the KL estimator (red dash-dot line). (b) Accumulation of mutual information as wavelengths are added. (c) Spectral library and first nine wavelengths selected by the three schemes. For (b) and (c), GROWLS used $p = 1$, $\delta = 1.6$.

3 Simulated Results and Discussion

In order to illustrate the utility of adaptive OWLS using GROWLS (“adaptive GROWLS”), we generated noisy synthetic data corresponding to liquid films on metallic substrates and assessed the chemical identification accuracy. As a simplification to the industry-standard adaptive cosine estimator (ACE),¹⁵ we used cosine similarity to match the simulated absorbance data a to the library signatures a_y :

$$CS(a, a_y) = \frac{a \cdot a_y}{||a|| ||a_y||}. \quad (5)$$

Figure 4(a) shows the contours corresponding to 90% chemical identification accuracy for evenly-spaced wavenumbers, *a priori* OWLS (i.e., wavelengths selected *a priori* using the KL estimator), and adaptive GROWLS. For this example, six wavelengths were added at each step and GROWLS used $q = 0.8$ (i.e., the top 20% of chemicals were retained at each iteration of Algorithm 1), $p = 1$, $\delta = 1.6$. The two OWLS schemes deliver comparable performance to each other, and require roughly half the number of wavelengths as the evenly-spaced scheme for moderate abundances.

The time to reach a detection of a given confidence is given by

$$t_d = N_b(N_\lambda t_m + t_c + t_s), \quad (6)$$

where t_d is the time-to-detection, N_b is the number of batches of N_λ wavelengths which are measured, t_m is the time to measure a single wavelength (~ 1 ms), t_c is the time to run the detection/identification algorithm to determine the detection confidence (and, for adaptive GROWLS, to rank the chemicals) after each batch (~ 10 μ s for cosine similarity with 67 candidates), and t_s is the time to select the next N_λ wavelengths (~ 2 ms for GROWLS with $N_\lambda = 6$). Figure 4(b) shows the time to reach a given level of accuracy for all three schemes. Despite the additional computational overhead, adaptive GROWLS delivers comparable per-

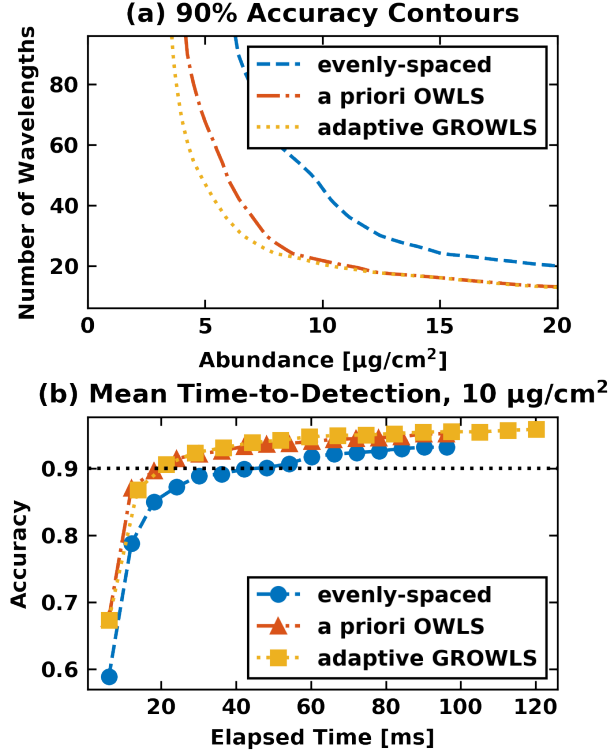


Figure 4: (a) Contours corresponding to 90% accuracy, averaged over all 67 chemicals in the library. Adaptive GROWLS delivers comparable average performance to *a priori* OWLS, and both OWLS schemes out-perform the evenly-spaced wavenumbers, thereby permitting successful detection at lower concentrations while using fewer wavelengths. (b) Time-to-detection when 10 $\mu\text{g}/\text{cm}^2$ of the target is present, averaged over all 67 chemicals. Despite the additional computational overhead, adaptive GROWLS delivers accuracy above 90% within about 20 ms, comparable to the performance of *a priori* OWLS.

formance to *a priori* OWLS: both OWLS schemes reach 90% accuracy in less than half the time it takes evenly-spaced wavenumbers to reach this point.

The comparable performance of *a priori* OWLS and adaptive GROWLS is expected at moderate concentrations: we are adding six wavelengths at each step, but often a few dozen wavelengths are sufficient at moderate concentrations, so adaptive GROWLS does not have much of a chance to make a difference. As noted previously, however, adaptive GROWLS makes a substantial difference for chemicals with low absorbance.²⁵ This is illustrated by Figure 5, which shows the detection performance when the target is 10 $\mu\text{g}/\text{cm}^2$ of benzene. Adaptive GROWLS obtains an accuracy of 90% after just 15 ms, compared to 30 ms for

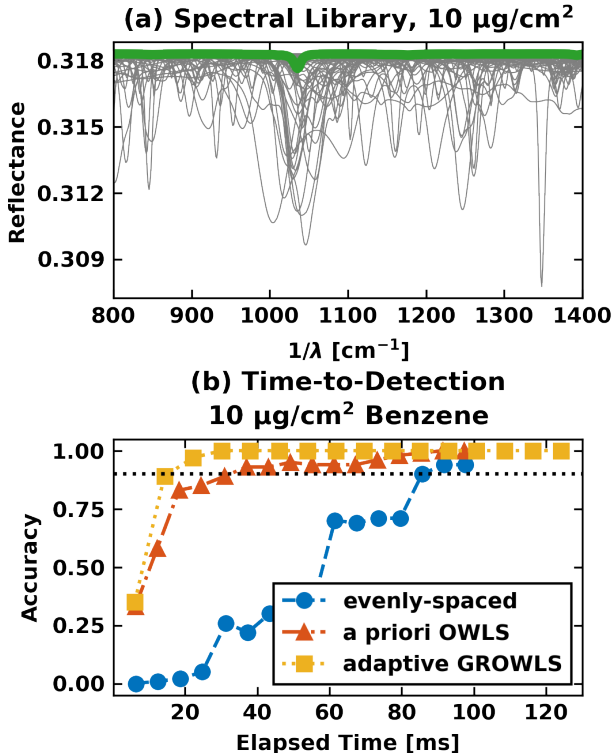


Figure 5: (a) Reflectance spectra for 10 $\mu\text{g}/\text{cm}^2$ of each chemical. Benzene (thick green) has very low absorbance with few strong spectral features compared to other chemicals, making it very challenging to identify. (b) Time-to-detection when the target is 10 $\mu\text{g}/\text{cm}^2$ of benzene.

a priori OWLS and 85 ms for evenly-spaced wavenumbers: adaptive GROWLS is able to substantially reduce the time-to-detection in this case.

4 Conclusions

This paper has presented a geometric approximation to mutual information which runs approximately $40\,000\times$ faster than the standard KL estimator, thereby enabling real-time adaptive selection of wavelengths to optimize detection accuracy. It was shown that, for moderate abundances, this approach halves the number wavelengths needed to reach an identification accuracy of 90%. For weak absorbers, the gains are even greater: benzene can be identified $6\times$ faster, even after accounting for the overhead to compute the next batch of wavelengths. These simulated results indicate that adaptive GROWLS can enable sensors with tunable sources to obtain more rapid detections. Furthermore, the approximation presented here is applicable to any situation where the classes have the general structure illustrated in Figure 2, enabling efficient feature selection in a wide variety of machine learning contexts.

Acknowledgement

The authors thank Dr. Kristin DeWitt, program manager of the IARPA SILMARILS program, for her support. Additionally, the authors thank their colleagues from Block MEMS lead by Dr. Anish Goyal for many useful discussions.

This research was funded by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the AFRL contract FA8650-16-C-9107. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI, or the U.S. Government. The Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

References

- (1) Larkin, P. J. *IR and Raman Spectroscopy*; Elsevier, 2011.
- (2) Hildenbrand, J.; Herbst, J.; Wöllenstein, J.; Lambrecht, A. Explosive detection using infrared laser spectroscopy. *Quantum Sensing and Nanophotonic Devices VI*. 2009; p 72220B.
- (3) Van Neste, C. W.; Senesac, L. R.; Thundat, T. Standoff Spectroscopy of Surface Adsorbed Chemicals. *Analytical Chemistry* **2009**, *81*, 1952–1956.
- (4) Pacheco-Londoño, L. C.; Otiz-Rivera, W.; Primera-Pedrozo, O. M.; Hernández-Rivera, S. P. Vibrational spectroscopy standoff detection of explosives. *Analytical and Bioanalytical Chemistry* **2009**, *395*, 323–335.
- (5) Wallin, S.; Pettersson, A.; Östmark, H.; Hobro, A. Laser-based standoff detection of explosives: a critical review. *Analytical and Bioanalytical Chemistry* **2009**, *395*, 259–274.
- (6) Goyal, A. K.; Spencer, M.; Kelly, M.; Costa, J.; DiLiberto, M.; Meyer, E.; Jeys, T. Active infrared multispectral imaging of chemicals on surfaces. *Chemical, Biological, Radiological, and Explosives (CBRNE) Sensing XII*. 2011; p 80180N.
- (7) Ortiz-Rivera, W.; Pacheco-Londoño, L. C.; Castro-Suarez, J. R.; Felix-Rivera, H.; Hernández-Rivera, S. P. Vibrational spectroscopy standoff detection of threat chemicals. *Micro- and Nanotechnology Sensors, Systems, and Applications III*. 2011; p 803129.
- (8) Macleod, N. A.; Rose, R.; Weidmann, D. Middle infrared active coherent laser spectrometer for standoff detection of chemicals. *Optics Letters* **2013**, *38*, 3708–3711.
- (9) Deutsch, E. R.; Kotidis, P.; Zhu, N.; Goyal, A. K.; Ye, J.; Mazurenko, A.; Norman, M.; Zafiriou, K.; Baier, M.; Connors, R. Active and passive infrared spectroscopy for the

- detection of environmental threats. Advanced Environmental, Chemical, and Biological Sensing Technologies XI. 2014; p 91060A.
- (10) Goyal, A. K.; Myers, T. R. In *Laser-Based Optical Detection of Explosives*; Pellegrino, P. M., Holthoff, E. L., Farrell, M. E., Eds.; CRC Press, 2015.
 - (11) Kotidis, P.; Deutsch, E. R.; Goyal, A. K. Standoff detection of chemical and biological threats using miniature widely tunable QCLs. Micro- and Nanotechnology Sensors, Systems, and Applications VII. 2015; p 94672S.
 - (12) Macleod, N. A.; Molero, F.; Weidmann, D. Broadband standoff detection of large molecules by mid-infrared active coherent laser spectrometry. *Optics Express* **2015**, *23*, 912–928.
 - (13) Kelley, D. B.; Wood, D.; Goyal, A. K.; Kotidis, P. High-speed and large-area scanning of surfaces for trace chemicals using wavelength-tunable quantum cascade lasers. Chemical, Biological, Radiological, and Explosives (CBRNE) Sensing XIX. 2018; p 1062909.
 - (14) Wavelength λ and wavenumber $1/\lambda$ are interchangeable. Most results here are presented in the conventional wavenumber units of cm^{-1} .
 - (15) Manolakis, D.; Shaw, G. Detection algorithms for hyperspectral imaging applications. *IEEE Signal Processing Magazine* **2002**, *19*, 29–43.
 - (16) Manolakis, D.; Lockwood, R.; Cooley, T.; Jacobson, J. Is there a best hyperspectral detection algorithm? Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XV. 2009; p 733402.
 - (17) Manolakis, D.; Truslow, E.; Pieper, M.; Cooley, T.; Brueggeman, M. Detection Algorithms in Hyperspectral Imaging Systems. *IEEE Signal Processing Magazine* **2014**, *31*, 24–33.

- (18) Raz, G.; Murphy, C.; Georgan, C.; Greenwood, R.; Prasanth, R. K.; Myers, T.; Goyal, A.; Kelley, D.; Wood, D.; Kotidis, P. Novel trace chemical detection algorithms: a comparative study. *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXIII*. 2017; p 101980D.
- (19) Murphy, K. P. *Machine Learning: A Probabilistic Perspective*; MIT Press, 2012.
- (20) Vergara, A.; Muezzinoglu, M. K.; Rulkov, N.; Huerta, R. Information-theoretic optimization of chemical sensors. *Sensors and Actuators B: Chemical* **2010**, *148*, 298–306.
- (21) Fonollosa, J.; Fernández, L.; Huerta, R.; Gutiérrez-Gálvez, A.; Marco, S. Temperature optimization of metal oxide sensor arrays using Mutual Information. *Sensors and Actuators B: Chemical* **2013**, *187*, 331–339.
- (22) Wang, X. R.; Lizier, J. T.; Nowotny, T.; Berna, A. Z.; Prokopenko, M.; Trowell, S. C. Feature Selection for Chemical Sensor Arrays Using Mutual Information. *PLoS ONE* **2014**, *9*, e89840.
- (23) Chang, C.-I. An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. *IEEE Transactions on Information Theory* **2000**, *46*, 1927–1932.
- (24) Fonollosa, J.; Vergara, A.; Huerta, R.; Marco, S. Estimation of the limit of detection using information theory measures. *Analytica Chimica Acta* **2014**, *810*.
- (25) Raz, G.; Chilenski, M.; Murphy, C.; Georgan, C. Optimal Sensor Control for Fast Target Detection in Hyperspectral Imagery. *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXIV*. 2018; p 1064407.
- (26) Cover, T. M.; Thomas, J. A. *Elements of Information Theory*; Wiley: Hoboken, New Jersey, 2006.

- (27) Fisher III, J. W.; Siracusa, M.; Tieu, K. Estimation of Signal Information Content for Classification. 2009 IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop. 2009; pp 353–358.
- (28) Kozachenko, L. F.; Leonenko, N. N. Sample Estimate of the Entropy of a Random Vector. *Probl. Peredachi Inf.* **1987**, *23*, 9–16.

Graphical TOC Entry

